In [123]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import confusion_matrix,classification_report
```

In [124]:

```
birds=pd.read_csv('/home/anudeep/Downloads/bird.csv')
birds.head()
```

Out[124]:

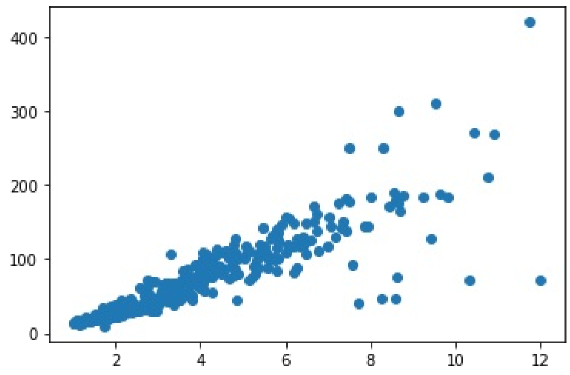
	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

In [125]:

```
plt.scatter(birds['ulnaw'],birds['huml'])
```

Out[125]:

```
<matplotlib.collections.PathCollection at 0x7f0c614c6890>
```

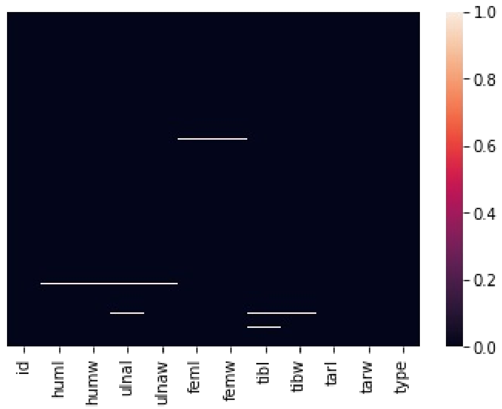


In [126]:

```
import seaborn as sns
sns.heatmap(birds.isnull(),yticklabels=False)
```

Out[126]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0c6147b610>
```



In [127]:

```
birds=birds.drop(['humw','ulnal','feml','femw','tibl','tibw','id','type','tarw','tarl'],axis=1)
birds.head()
```

Out[127]:

	huml	ulnaw
0	80.78	4.88
1	88.91	5.59
2	79.97	5.28
3	77.65	4.77
4	62.80	3.73

In [128]:

```
birds.describe()
```

Out[128]:

	huml	ulnaw
count	419.000000	418.000000
mean	64.650501	3.597249
std	53.834549	2.186747
min	9.850000	1.000000
25%	25.170000	1.870000
50%	44.180000	2.945000
75%	90.310000	4.770000
max	420.000000	12.000000

In [129]:

```
birds.isnull()
```

Out[129]:

	huml	ulnaw
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...	...	...
415	False	False
416	False	False
417	False	False
418	False	False
419	False	False

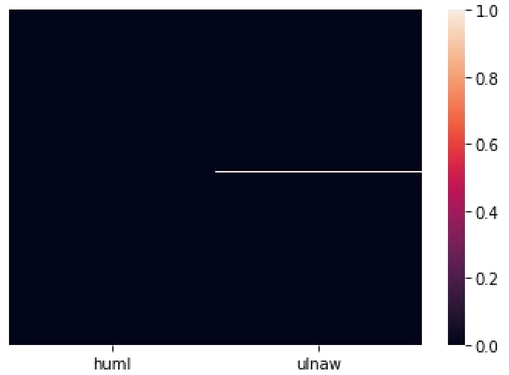
420 rows × 2 columns

In [130]:

```
import seaborn as sns
sns.heatmap(birds.isnull(),yticklabels=False)
```

Out[130]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0c6142c8d0>
```



In [131]:

```
birds=birds.dropna() # removes rows having any single null value
sns.heatmap(birds.isnull(),yticklabels=False)
```

Out[131]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0c613a1510>
```

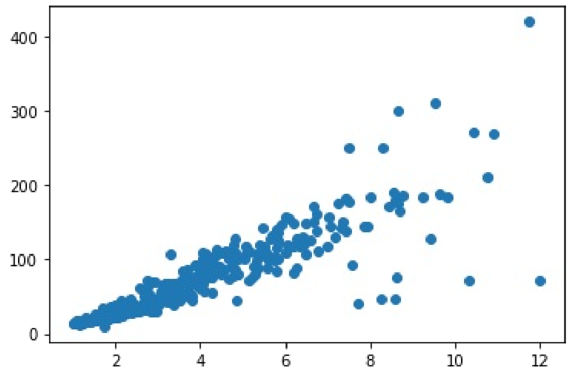
0.08
0.04
0.00
−0.04
−0.08
huml
ulnaw

In [132]:

```
plt.scatter(birds['ulnaw'],birds['huml'])
```

Out[132]:

```
<matplotlib.collections.PathCollection at 0x7f0c612f0e10>
```



In [133]:

```
km=KMeans(n_clusters=3)
km
```

Out[133]:

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)
```

In [134]:

```
y_predicted=km.fit_predict(birds[['huml','ulnaw']])
y_predicted
```

Out[134]:

```
array([1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 0, 0, 0, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1,
       1, 1, 1, 2, 2, 1, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 2, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 1, 1, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0,
       0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
      dtype=int32)
```

In [136]:

```
scaler = MinMaxScaler()

birds = pd.DataFrame(scaler.fit_transform(birds), columns=birds.columns)

birds
```

Out[136]:

	huml	ulnaw
0	0.172937	0.352727
1	0.192759	0.417273
2	0.170962	0.389091
3	0.165305	0.342727
4	0.129099	0.248182
...	...	...
413	0.019773	0.030000
414	0.022821	0.044545
415	0.021797	0.048182
416	0.025674	0.045455
417	0.019603	0.009091

418 rows × 2 columns

In [137]:

```
birds['cluster']=y_predicted
birds.head()
```

Out[137]:

	huml	ulnaw	cluster
0	0.172937	0.352727	1
1	0.192759	0.417273	1
2	0.170962	0.389091	1
3	0.165305	0.342727	1
4	0.129099	0.248182	0

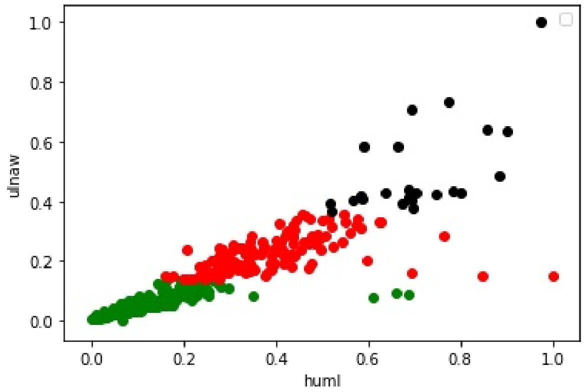
In [138]:

```
df1=birds[birds.cluster==0]
df2=birds[birds.cluster==1]                # 5 different data frames each belonging to one cluster
df3=birds[birds.cluster==2]
#df4=birds[birds.cluster==3]
#df5=birds[birds.cluster==4]
#df6=birds[birds.cluster==5]

plt.scatter(df1.ulnaw,df1['huml'],color='green')
plt.scatter(df2.ulnaw,df2['huml'],color='red')
plt.scatter(df3.ulnaw,df3['huml'],color='black')

plt.xlabel('huml')
plt.ylabel('ulnaw')
plt.legend()
```

```
No handles with labels found to put in legend.
```

Out[138]:

```
<matplotlib.legend.Legend at 0x7f0c6120b7d0>
```



In [139]:

```
birds.head()
```

Out[139]:

	huml	ulnaw	cluster
0	0.172937	0.352727	1
1	0.192759	0.417273	1
2	0.170962	0.389091	1
3	0.165305	0.342727	1
4	0.129099	0.248182	0

In [140]:

```
km=KMeans(n_clusters=3)
y_predicted=km.fit_predict(birds[['huml','ulnaw']])
y_predicted
```

Out[140]:

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1,
       1, 1, 1, 2, 2, 2, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 2, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 1, 2, 2, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 2, 1, 1, 1, 2,
       1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 2, 1,
       0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2,
       1, 0, 0, 1, 2, 0, 1, 1, 1, 0, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1,
       2, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 2, 2, 2, 0, 0, 1, 1, 1, 1, 1, 0,
       0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
      dtype=int32)
```

In [141]:

```
birds['cluster1']=y_predicted
birds.drop('cluster',axis='columns',inplace=True)
birds
```

Out[141]:

	huml	ulnaw	cluster1
0	0.172937	0.352727	1
1	0.192759	0.417273	1
2	0.170962	0.389091	1
3	0.165305	0.342727	1
4	0.129099	0.248182	1
...	...	...	...
413	0.019773	0.030000	0
414	0.022821	0.044545	0
415	0.021797	0.048182	0
416	0.025674	0.045455	0
417	0.019603	0.009091	0

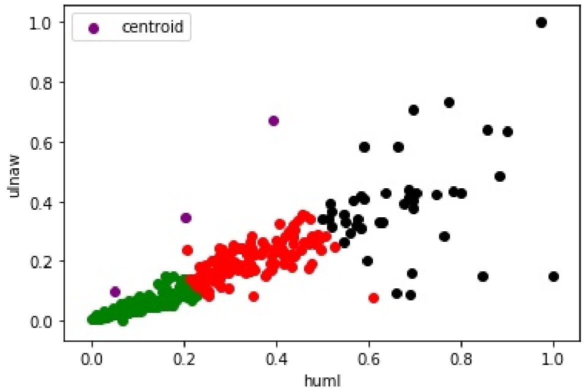
418 rows × 3 columns

In [148]:

```
df1=birds[birds.cluster1==0]
df2=birds[birds.cluster1==1]             # 5 different data frames each belonging to one cluster
df3=birds[birds.cluster1==2]
#df4=birds[birds.cluster==3]
#df5=birds[birds.cluster==4]
#df6=birds[birds.cluster==5]

plt.scatter(df1.ulnaw,df1['huml'],color='green')
plt.scatter(df2.ulnaw,df2['huml'],color='red')
plt.scatter(df3.ulnaw,df3['huml'],color='black')
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',label='centroid')

plt.xlabel('huml')
plt.ylabel('ulnaw')
plt.legend()
```

Out[148]:

```
<matplotlib.legend.Legend at 0x7f0c611c1c90>
```



In [146]:

```
km.cluster_centers_
```

Out[146]:

```
array([[0.04921461, 0.10011645],
       [0.2045302 , 0.34515604],
       [0.39367943, 0.67181818]])
```

In [149]:

```
k_rng =range(1,10)
sse=[]
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit_predict(birds[['huml','ulnaw']])
    sse.append(km.inertia_)
```

In [150]:

```
sse
```

Out[150]:

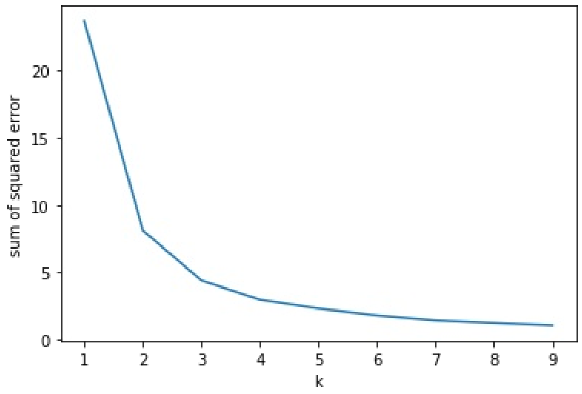
```
[23.68097920799888,
 8.114806978881068,
 4.4002670800893675,
 2.9674484389170788,
 2.311481571446346,
 1.7951443360064623,
 1.4252434557499394,
 1.229590458446783,
 1.0516941849447923]
```

In [151]:

```
plt.xlabel('k')
plt.ylabel('sum of squared error')
plt.plot(k_rng,sse)
```

Out[151]:

```
[<matplotlib.lines.Line2D at 0x7f0c61153190>]
```



In [ ]: